

## 2.2 CPU的逻辑结构与工作原理

**2.2.1 CPU的逻辑结构**

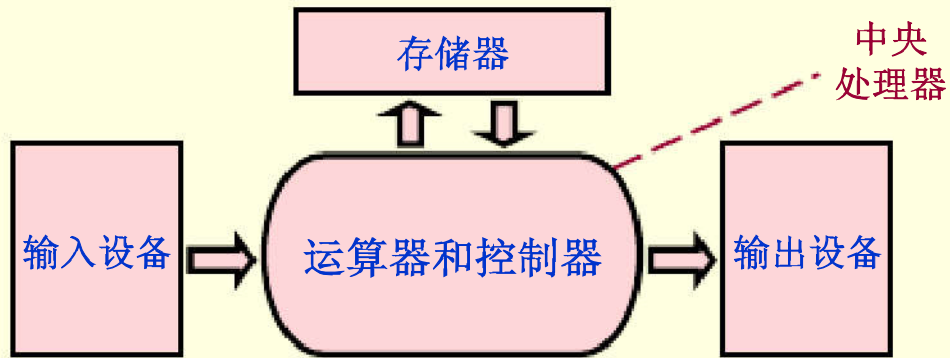
**2.2.2 指令与指令系统**

**2.2.3 CPU的性能指标**

---

## 2.2.1 CPU的逻辑结构

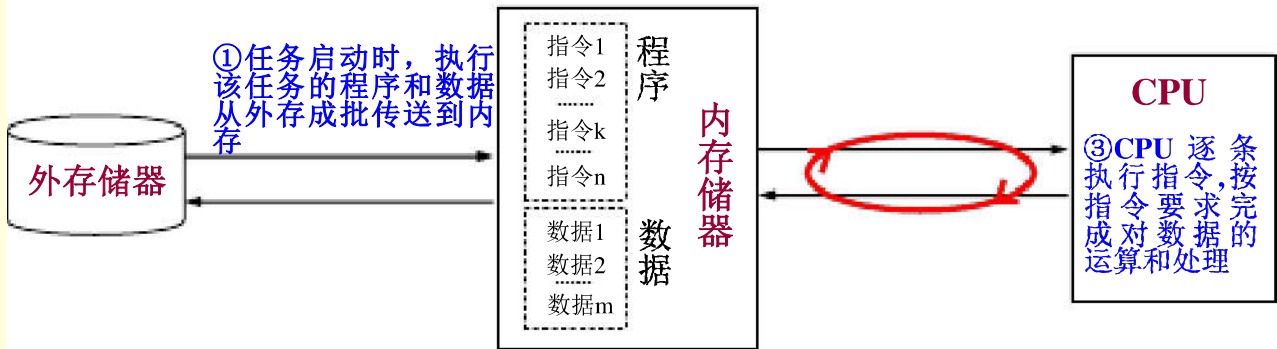
# 冯·诺依曼计算机的结构与原理



- (1) 计算机的工作由程序控制，程序是一个指令序列，指令是能被计算机理解和执行的操作命令；
- (2) 程序(指令)和数据均以二进制编码表示，均存放在存储器中；
- (3) 存储器中存放的指令和数据按地址进行存取；
- (4) 指令是由CPU一条一条顺序执行的。

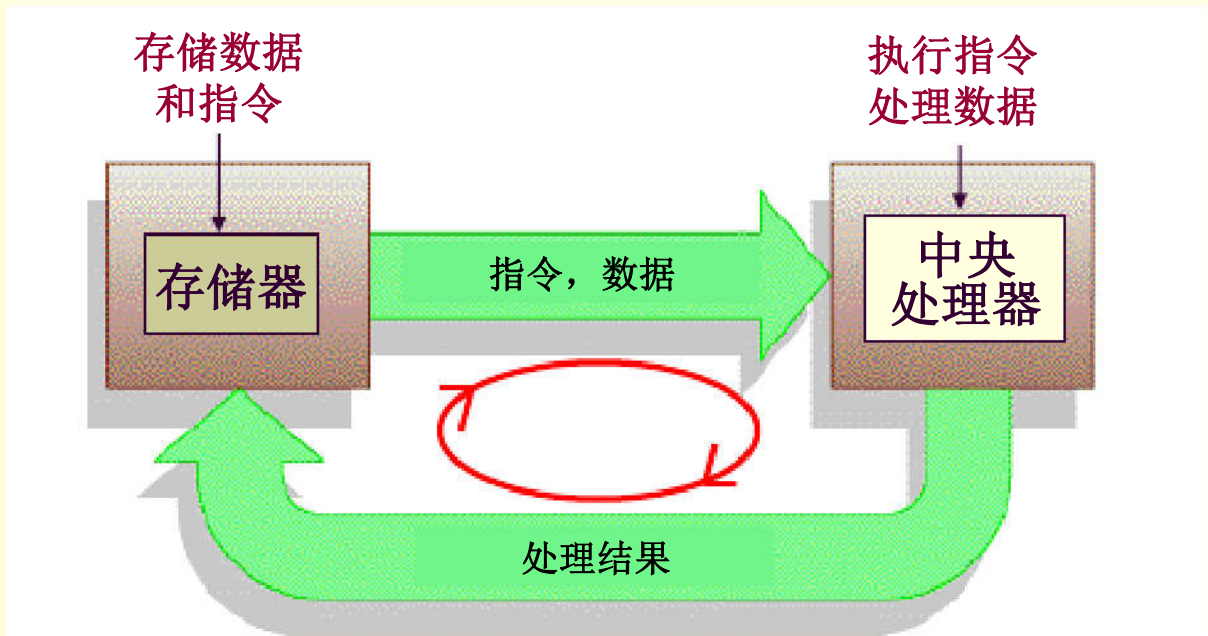
# “存储程序控制” 原理

- 将问题的解算步骤编制成为程序，程序连同它所处理的数



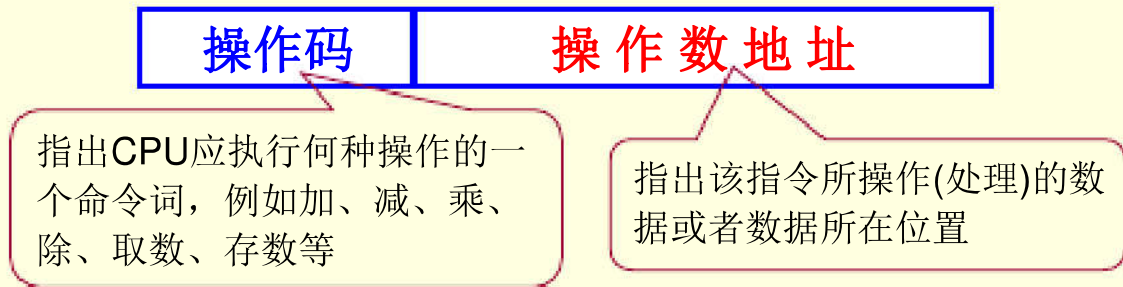
# CPU的任务

- CPU的主要任务是执行指令，它按指令的规定对数据进行操作



# 指令是什么？

- 指令就是命令，它用来规定CPU执行什么操作。指令是构成程序的基本单位，程序是由一连串指令组成的
- 指令采用二进位表示，大多数情况下，指令由两个部分组成：



举例: **10 02 06**

把**02**存储单元和**06**存储单元中的内容相加，和数保存在**02**单元

# CPU的结构和任务

- CPU主要由运算器、控制器和寄存器组3部分组成
- CPU的任务：取指令并完成指令所规定的操作



# 运算器(ALU)与通用寄存器(GPR)

- 运算器用来对数据进行各种算术或逻辑运算，所以称为算术逻辑部件 (ALU)，参加ALU运算的操作数通常来自通用寄存器 GPR，运算结果也送回GPR

## 例1:取数指令

LOAD	R3	内存地址A
LOAD	R5	内存地址B

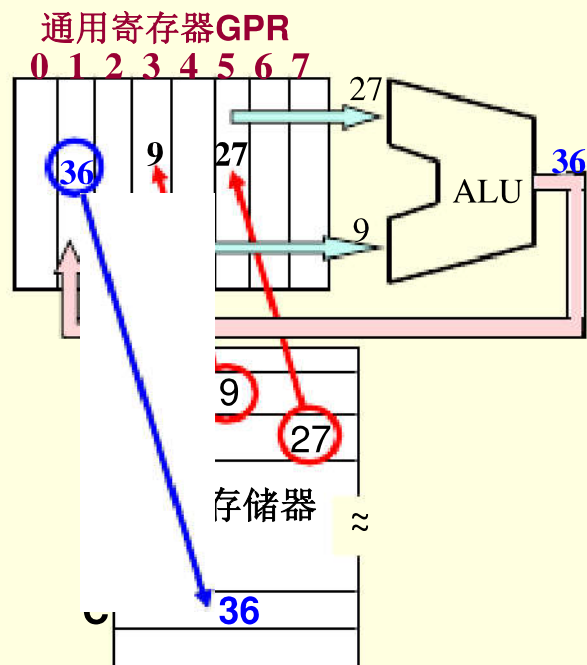
## 例2: 加法指令

ADD	R1	R3	R5
-----	----	----	----

(3# 寄存器内容与5# 寄存器内容相加，并把和数写入1# 寄存器)

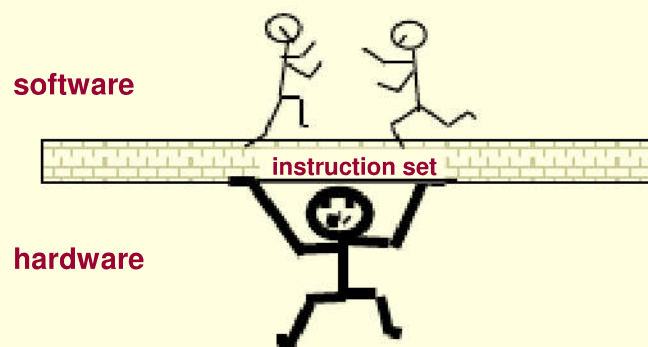
## 例3: 存数指令

STORE	R1	内存地址C
-------	----	-------



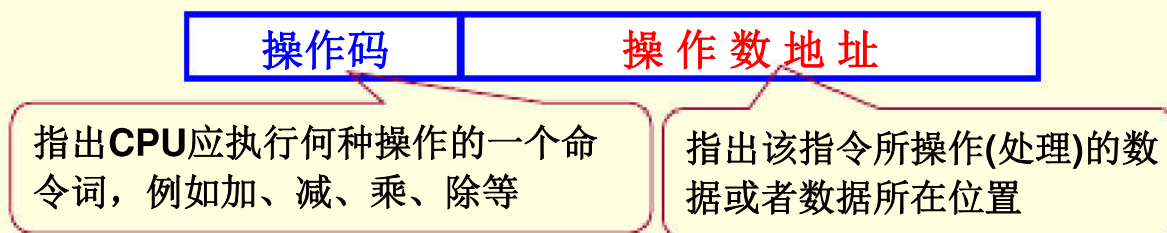


## 2.2.2 指令与指令系统



# 指令及其格式

- 指令采用二进制表示，大多数情况下由两个部分组成：



## 指令的格式：

- 操作码用几位表示？(多少种不同的操作功能)
- 操作数地址是2个还是3个？

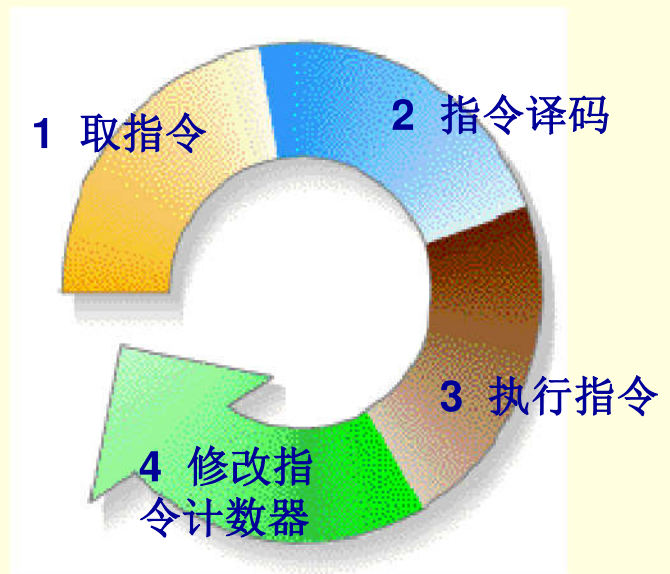
■ $R1 \leftarrow R3 + R5$	ADD	R1	R3	R5
■ $R1 \leftarrow R1 - R3$	SUB	R1	R3	

- 如何指出内存地址？(寻址方式)
- 指令长度是多少位，固定长度/可变长度？

# 指令在计算机中的执行过程

1. **取指令**：CPU的控制器从存储器读取一条指令并放入指令寄存器
2. **指令译码**：指令寄存器中的指令经过译码，决定该指令应进行何种操作、操作数在哪里
3. **执行指令**
  - 3.1 取操作数
  - 3.2 进行运算
4. **修改指令计数器**，决定下一条指令的地址

## 指令执行周期

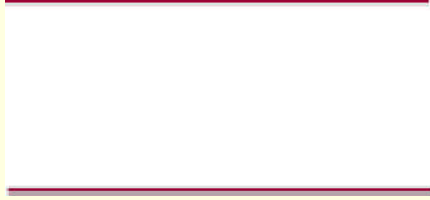


问题：CPU的时钟频率(主频)与指令执行速度是什么关系？

# 指令执行过程的模拟

---

模拟程序1 (flash动画, 单步)



(flash动画, 自动)

# 关于指令系统

- **CPU**可执行的全部指令称为该**CPU**的指令系统，即它的机器语言
- 指令应该简单，简单才能提高速度！
  1. 操作功能要简单，不宜太多、太复杂
  2. 指令的格式要统一：固定长度、统一分段
  3. 操作数的来源要有限制：
    - 算术逻辑运算指令的操作数只能来自寄存器
    - **LOAD/STORE**指令的操作数来自(或送到)内存
      - 寄存器←内存      内存←寄存器
- **Intel 奔腾**处理器的指令系统由于历史的原因，并不符合上述原则！

# 关于指令的操作功能

- 指令系统中的指令分成许多类，例如奔腾4处理器中共有七大类指令：
  - 数据传送类
  - 算术运算类
  - 逻辑运算类
  - 移位操作类
  - 位(位串)操作类
  - 控制转移类
  - 输入/输出类等
- 每一类指令（如数据传送类、算术运算类）又按照操作数的性质（如整数还是实数）、长度（16位、32位、64位、128位等）而区分为许多不同的指令，因此CPU往往有数以百计的不同的指令

# 关于奔腾4系列的指令系统

- 产品发展过程为：**8088（8086）→80286→80386→80486→Pentium→Pentium PRO→Pentium II→Pentium III→Pentium 4→奔腾D→奔腾至尊→酷睿→酷睿2**
- 为解决软件兼容性问题，采用“向下兼容方式”开发新的处理器，即所有新处理器均保留老处理器的全部指令，同时还扩充功能更强的新指令。例如：
  - Pentium比80386增加了80多条浮点指令，
  - Pentium II比Pentium增加了50多条处理多媒体信息的指令（**Multimedia Extension**，称为**MMX**），
  - Pentium III比Pentium II增加了70条处理128位长操作数的流式单指令多数据指令（**Streaming SIMD Extention**，**SSE**）
  - Pentium 4又比Pentium III增加了144条指令（称为**SSE2**）
  - 奔腾D、奔腾至尊等又增加了**SSE3**指令

# 关于CPU的兼容性

- 不同公司生产不同的**CPU**产品：
  - Intel公司：Pentium系列、赛扬系列、至强等
  - AMD公司：独龙、闪龙、炫龙等
  - IBM公司：Power5、Power6等（苹果的Macintosh使用）
  - SUN公司：Ultra SPARC处理器
  - HP公司：PA-RISC、Itanium 处理器
- 同一公司同一系列的**CPU**具有向下（前）兼容性：
  - A计算机（使用新型号**CPU**）的程序不能在B计算机（使用老型号**CPU**）上运行，B计算机的程序却能在A计算机上运行
- 不同公司的不同**CPU**产品其指令系统不同，它们互相不兼容：
  - A计算机的程序不能在B计算机上直接运行，B计算机的程序也不能在A计算机上直接运行

互相兼容



---

## 2.2.3 CPU的性能指标

# 如何衡量CPU的性能？

- 计算机的性能主要表现在程序执行速度的快慢，它由许多因素决定，例如**CPU**、内存、硬盘、显卡等，但通常**CPU是主要因素**
- **CPU性能高低的主要指标是CPU的速度**，有2种衡量方法：
  1. 计算每秒钟可执行的指令数目（单位：**MIPS**、**MFLOPS**）
  2. 使用常用软件（办公软件、数字媒体处理软件和**3D**游戏等）的执行速度来衡量**CPU**的性能

# 影响CPU性能的主要因素(之一)

- **CPU的字长（位数）**
  - 目前**PC**使用的**CPU**大多是**32**位处理器，新一代的**PC**机将使用**64**位处理器
- **主频（CPU时钟频率）**
  - 主频提高，**CPU**的处理速度通常也会加快
- **CPU总线（前端总线）的速度**
  - **CPU**总线速度决定了**CPU**与内存间数据传输速度的快慢
- **高速缓存（cache）的容量与结构**
  - **cache**容量越大、级数越多，其效用就越显著

# 影响CPU性能的主要因素(之二)

## ■ CPU的指令系统

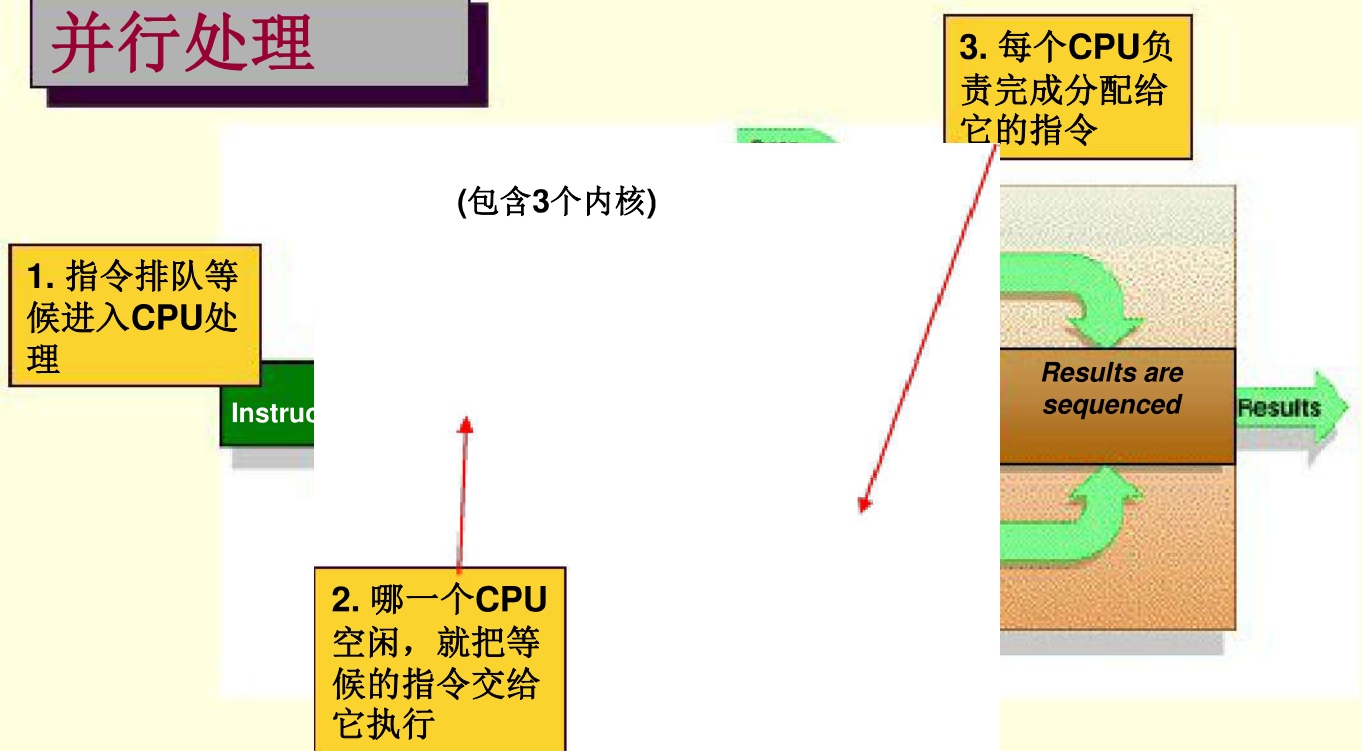
- 指令的格式和功能会影响程序的执行速度

## ■ CPU的逻辑结构

- CPU包含的定点运算器和浮点运算器数目
- 是否流水线结构，流水线的条数和级数
- 有无指令预测和数据预测功能
- 是否具有数字信号处理功能
- 是否多核，有几个内核
- .....

# 提高计算机速度的有效途经

## 并行处理



# Intel 微处理器主要技术参数比较

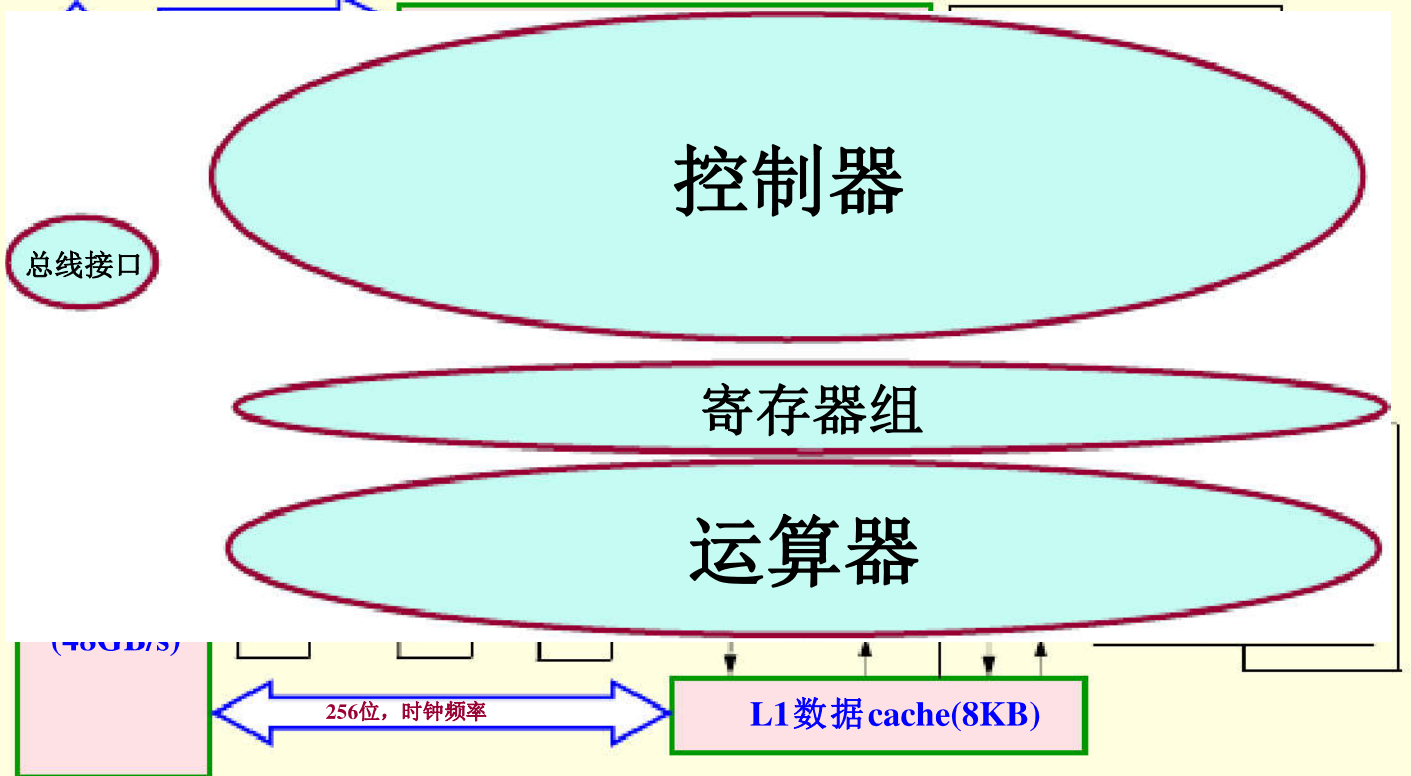
处理器主要参数	8080	8086	80286	80386	80486	奔腾	高能奔腾	奔腾II	奔腾III	奔腾4
推出时间(年)	1974	1978	1982	1985	1989	1993-96	1995-97	1997-98	1999-2003	2000-2006
主频 (MHz)	2	4.77	6-20	16-33	33-100	60-200	150-200	233-333	450-1400	1500-3800
前端总线频率 (MHz)	2	4.77	6~20	16~33	25或33	50或66	66	66	100或133	400, 533或800, 1066
外部数据线数目	8	16	16	32	32	64	64	64	64	64
地址线数目	16	20	24	32	32	32	36	36	36	36
存储器空间大小	64K	1MB	16MB	4GB	4GB	4GB	64GB	64GB	64GB	64GB
晶体管数目(万)	0.45	2.9	13.4	27.5	120	310	550	750	950	4200
制造工艺( $\mu\text{m}$ )	6	>2	1.5	1.5~1.0	1.0~0.8	0.8~0.35	0.6-0.35	0.35-0.25	0.25-0.13	0.13~0.065
芯片引脚数目	40	,40	68	132	168	273或296	387	242	370	478或775

# 附：Pentium4的逻辑结构

---

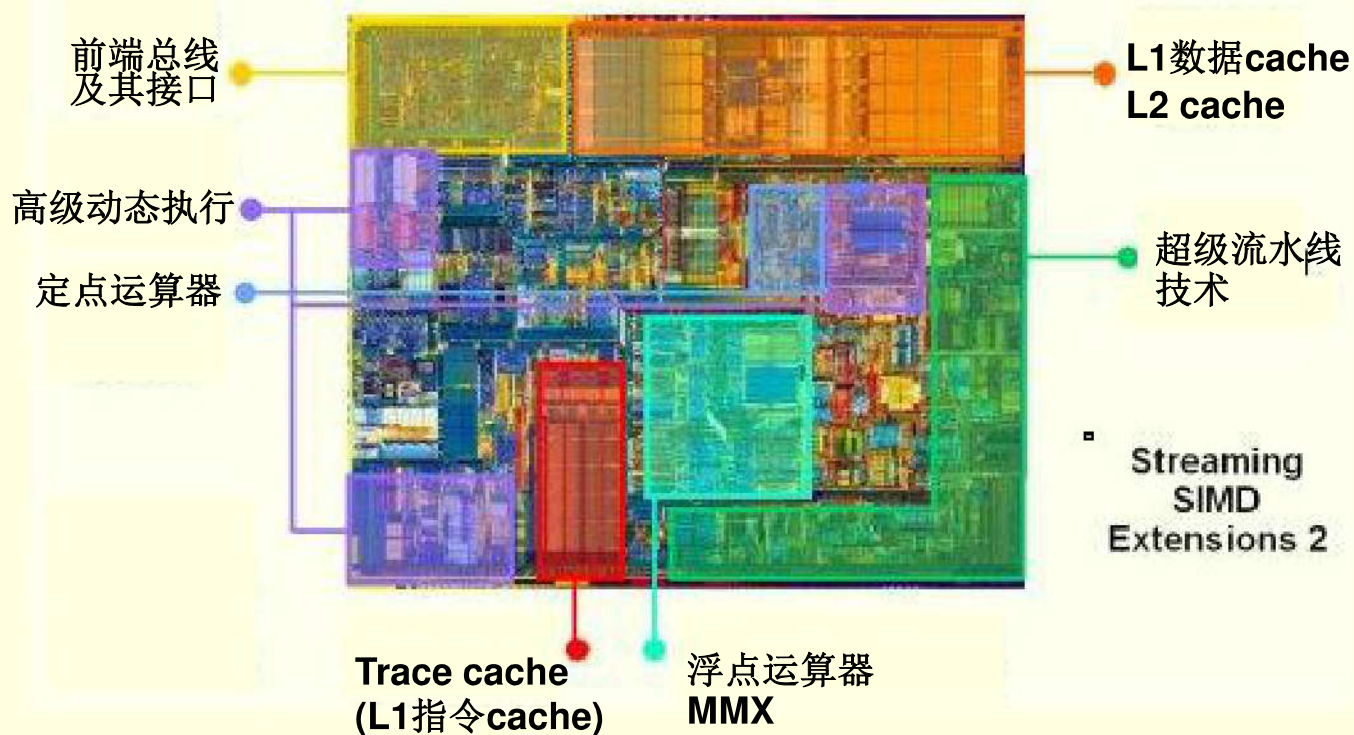
- 寄存器组与CPU字长
- 超标量结构ALU
- 流水线处理技术
- SIMD技术
- EM-64T
- 超线程技术
- 双核与多核技术

# Pentium 4 处理器的逻辑结构



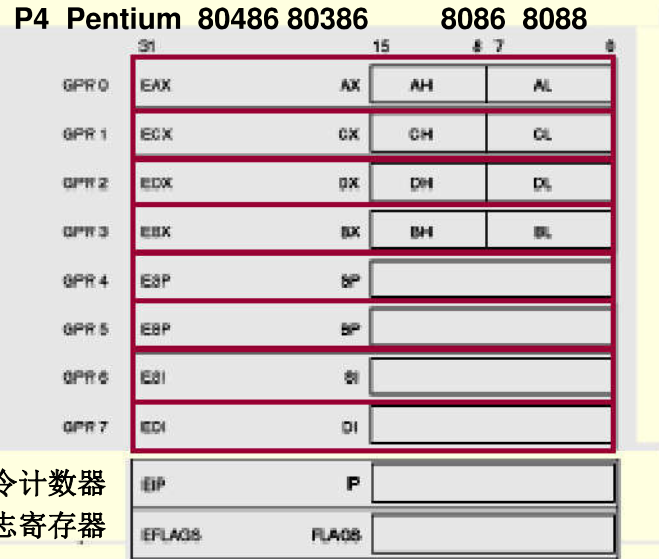


# Pentium 4 处理器的芯片布局

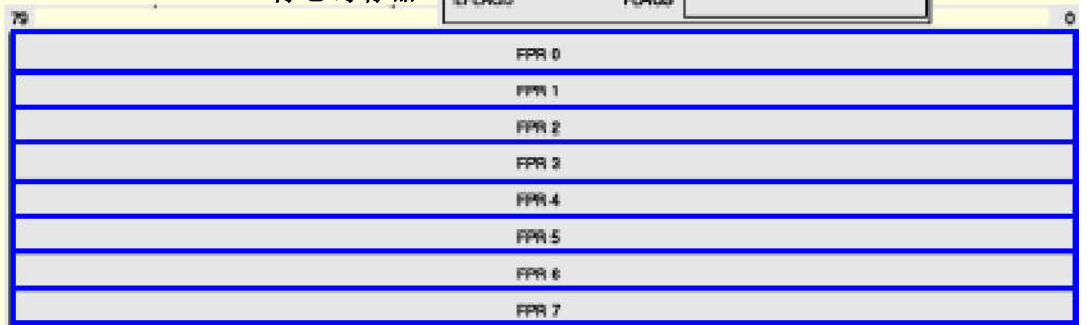


# Pentium 4 的寄存器组

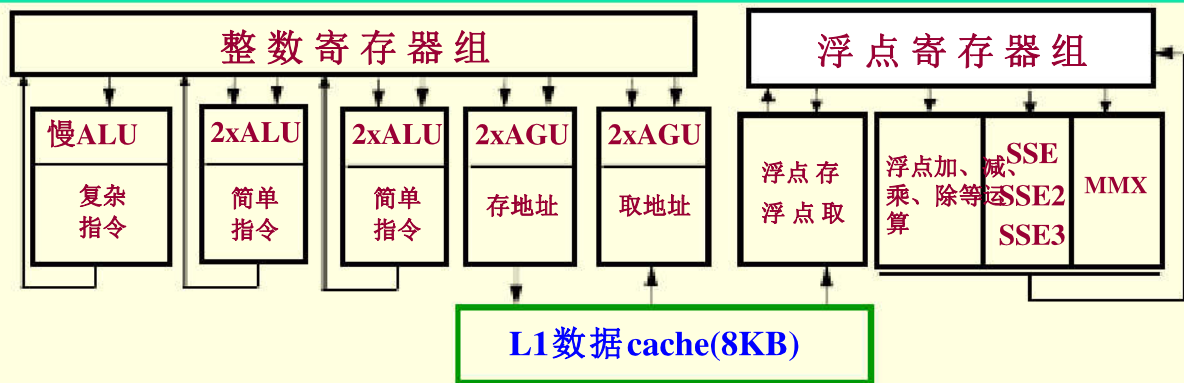
## 整数寄存器组



## 浮点寄存器组



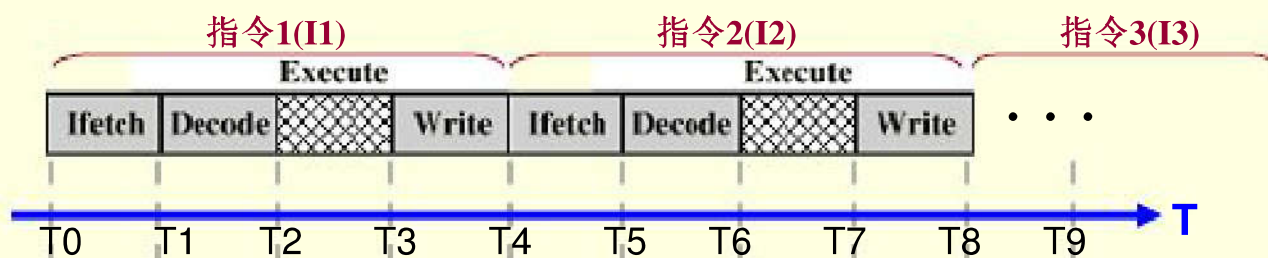
# Pentium 4的超标量结构运算器



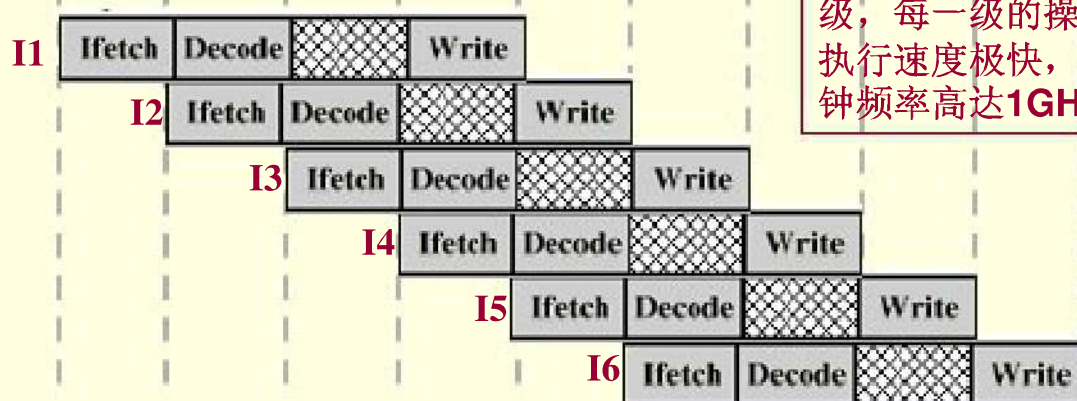
- 采用超标量（**superscalar**）结构，一共包含**9个ALU**，均可同时工作：
  - 2个高速整数ALU(每个时钟周期进行2次操作)，用于完成简单的整数运算(如加、减法)
  - 1个慢速整数ALU(需要多个时钟周期才能完成1次操作)，用于完成整数乘、除法运算
  - 2个地址生成部件（AGU），用于计算操作数的有效地址，所生成的地址分别用于从内存取操作数或向内存保存操作结果
  - 1个ALU用于完成浮点操作数地址的计算
  - 1个ALU用于完成浮点加法、乘法和除法运算
  - 1个ALU用于执行流式的SIMD处理（SSE/SSE2/SSE3指令）
  - 1个ALU用于完成多媒体信号处理（MMX指令）

# 指令的流水线执行

指令的顺序执行：



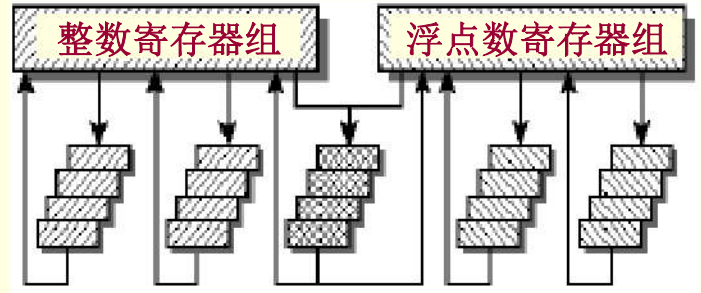
指令的流水线执行：



**Pentium 4**的流水线分成**20**级，每一级的操作都很简单，执行速度极快，因而允许时钟频率高达**1GHz**以上

# 超流水线(Hyper-pipeline)技术

- Pentium 4有多条超流水线，每条流水线的级数均很长，定点运算达20级，浮点运算达到29级，处于执行状态的指令数最多可达到126条



- 超级流水线中每步操作都非常简单，因此主频可以显著提高

- Pentium 4 的20级超流水线结构的示意图：



取指阶段

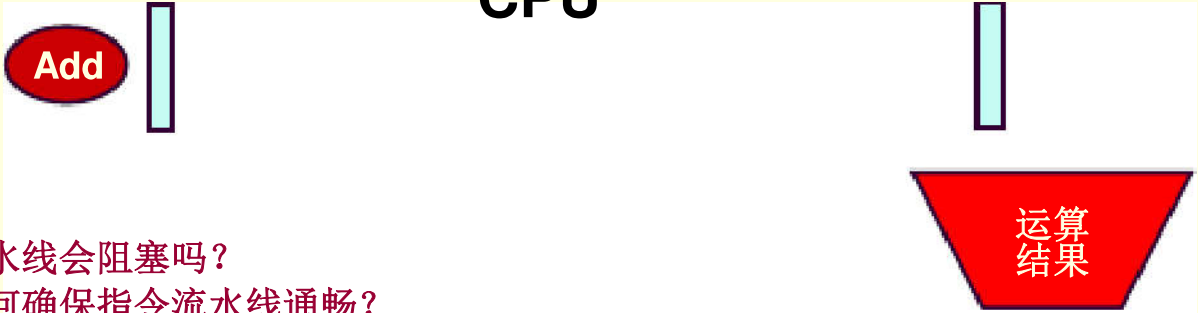
译码阶段

执行阶段

保存结果,修改IP

# 指令流水线的效果

CPU



流水线会阻塞吗？  
如何确保指令流水线通畅？

例：

$a = b + c;$   
 $e = e - d;$

假设a,b,c,d,e分别存储在地址为A,B,C,D,E的单元中



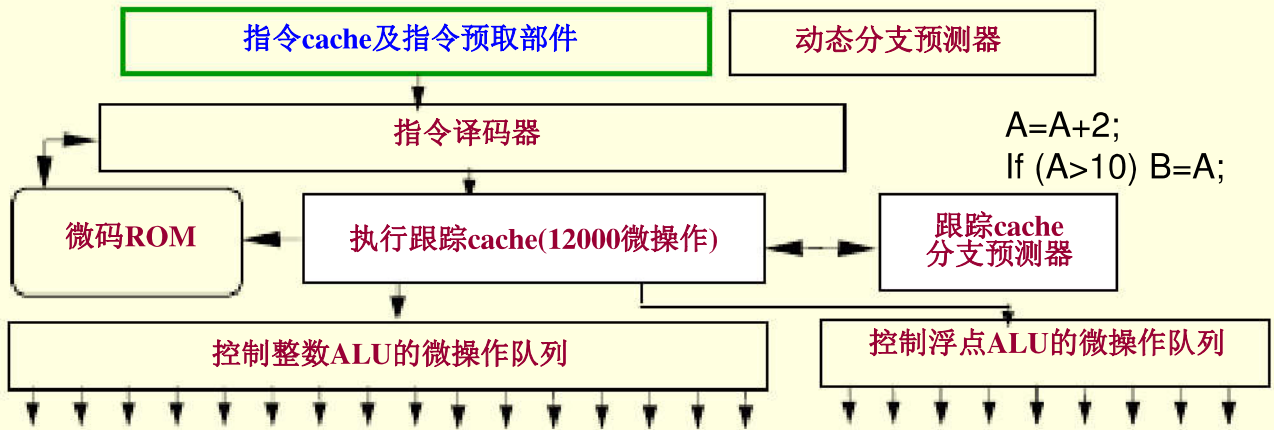
LOAD R1, B  
LOAD R2, C  
• ADD R2, R1  
• STORE R2, A  
LOAD R3, D  
LOAD R4, E  
• SUB R4, R3

调整后：

LOAD R1, B  
LOAD R2, C  
LOAD R3, D  
ADD R2, R1  
LOAD R4, E  
STORE R2, A  
SUB R4, R3

# 指令预取和分支预测

- 为了使流水线不中断，指令预取部件用于完成指令地址的计算，并从指令 **cache**中读取指令（一次**64**位），它还通过动态分支预测器对即将执行的指令提前进行预取，如果预测发生错误，那么流水线就会中断，**CPU**的速度将会受到影响



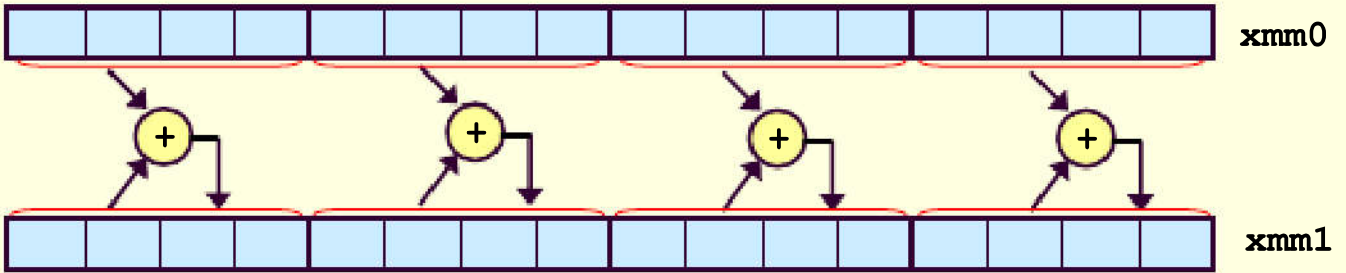
# MMX/SSE/SSE2/SSE3指令

- 所谓**SSE**指令，其特点是1条指令可处理**128位**的数据，它可以是下列情况之一：
  - 4个单精度浮点数(各**32位**)
  - 2个双精度浮点数(各**64位**)
  - 16个**8位**整数
  - 8个**16位**整数
  - 4个**32位**整数
  - 2个**64位**整数
  - 1个**128**整数
- 由于一条指令可以处理若干个整数或浮点数，因而大大提高了计算速度，这种做法称为单指令多数据技术（**single-instruction, multiple-data**，简称**SIMD**）
- **SSE/SSE2/SSE3**指令在完成**3D**图形、语音识别、图像处理等多媒体应用的时候非常有效

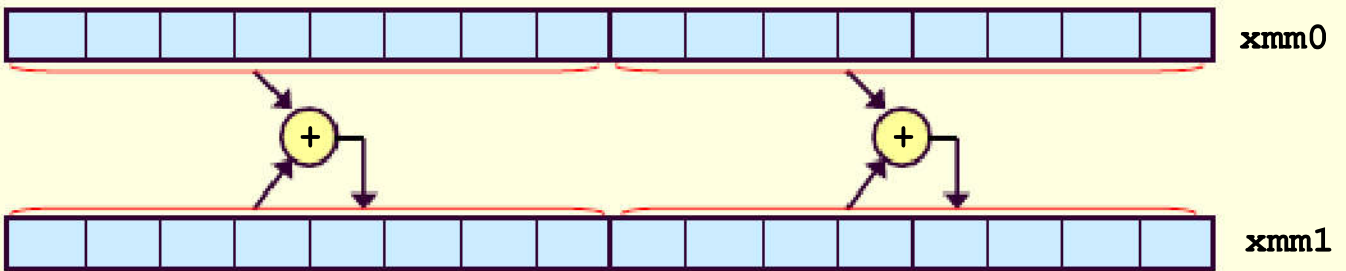


# SSE指令的SIMD操作举例

指令: `addps xmm0, xmm1` (1次完成4个单精度浮点数相加)



指令: `addpd xmm0, xmm1` (1次完成2个双精度浮点数相加)



# 为什么需要64位计算？

- 原因：
  - 可以进行更大范围的整数运算
  - 可以支持更大的内存
- 进行64位计算的条件：
  - 64位的处理器
  - 64位的操作系统(如 Windows XP X64, Windows Vista)
  - 64位的应用软件
- 产品：
  - RISC(UltraSparc III、IBM公司的POWER5、HP公司的Alpha)
  - Intel 安腾处理器( Itanium 和 Itanium 2)
  - AMD64(Athlon 64系列、Athlon FX系列和Opteron系列)
  - Intel EM-64T(Extended Memory 64 Technology)
    - 至强(Xeon)系列、Pentium 4 6xx系列 和 Pentium D系列、P4 至尊版

# Pentium的64位扩展技术(EM-64T)

## Pentium 4 的64位寄存器组

整数寄存器组

P4 Pentium 80386 80486 8086 8088

64位处理器的整数寄存器组



- 共16个64位寄存器，另外还增加8个128位的SSE寄存器
- 增加了处理64位整数的指令
- 支持c语言中的“long int”数据类型，对应的是64位整数

# 超线程(Hyper-Threading, HT)技术

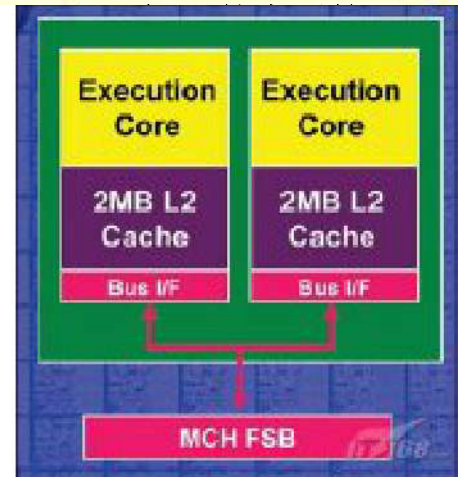
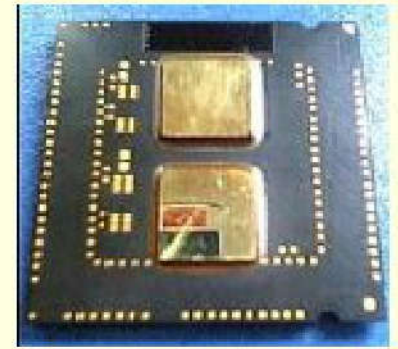
- 背景：
  - 实际应用中**CPU**的执行单元没有被充分使用，性能未得到充分发挥
- 超线程(HT)技术：把硬件模拟成两个处理器芯片，从**OS**来看就像有两个**CPU**一样，可同时执行2个线程
  - **P4**处理器增加了一个逻辑**CPU**指针，而整数运算单元、浮点运算单元、**L2 Cache**等均由2个线程共享
  - 处理器有两种运行模式：
    - **Single Task Mode**（单任务模式）：系统只有1个逻辑处理器
    - **Multi Task Mode**（多任务模式）：系统有2个逻辑处理器
- 分析：
  - 由于不是2个真正的**CPU**，它们需要共享**ALU**、**cache**等资源，当两个线程同时需要使用某个资源时，一个线程必须暂停运行，直到该资源空闲后才能继续执行。因此超线程的性能达不到2个物理**CPU**的性能
  - 超线程技术需要芯片组、**OS**和应用软件支持，才能发挥该项技术的优势，**Windows2000**就不支持双线程
  - 当运行单线程应用软件时，超线程技术甚至会降低系统性能

# 双核处理器出现的背景

- 提高主频来提升处理器性能的瓶颈是散热问题：
  - 3.2GHz 的Pentium 4处理器功率超过100W，内核温度达摄氏70度
  - 提升到4.0GHz时功率会达到150W，散热问题更难处理
- 超线程技术虽然可以提高执行部件的使用效率，但有一定开销，对于单线程的软件反而降低了效率
- 集成电路制造及封装技术的进步，有能力把2个甚至更多个处理器做在1个芯片内

# 双核处理器

- 1个芯片中有两个功能相同的处理器(内核), 在操作系统看来, 系统中有2个物理处理器
- 2个内核可以各有自己的L2 cache, 但必须保证其中的信息完全一致, 否则就会出现运算错误 (由955X芯片组中北桥芯片负责), 也可以共享同一个L2 cache
- 用途: 面向计算密集型应用和娱乐发烧友



独立  
cache

共享  
cache

2.2 CPU的逻辑结构与原理

# 小结：Pentium4提高速度的措施

- 扩展CPU的字长：64位存储扩展
- 提高CPU的主频：已经从1.5GHz逐步提高到3GHz以上，目前最高已经接近4GHz；
  - 加快CPU前端总线的数据传输速度：
    - CPU总线宽度增加为64~128位
    - CPU总线频率从400MHz、533 MHz提高到800MHz和1066MHz
    - 因此，传输速率也相应地从3.2GB/s、4.3GB/s提高到6.4GB/s和8.6GB/s
- 采用cache存储器
  - 增大cache容量：L2 cache的容量从256KB也已经增大为1MB或2MB
  - 增加cache的级数：有些处理器采用L3 cache
- 采用超标量运算器结构和超流水线技术
- 提供和支持向量运算指令（SIMD指令）
- 采用超线程技术，提高执行部件的工作效率
- 采用双(多)核处理器技术